DOCUMENT RESUME

ED 247 307 TM 840 490

AUTHOR Jones, Patricia B.; Sabers, Darrell L.
TITLE An Investigation of Two Procedures for Smoothing Test Norms.
PUB DATE Apr 84
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Equated Scores; Goodness of Fit; Measurement Techniques; Norm Referenced Tests; Scoring; *Statistical Analysis; *Statistical Distributions; Test Construction; Testing Problems; *Test Norms; Test Theory
IDENTIFIERS *Smoothing Methods; *Spline Function

ABSTRACT

Several techniques have been developed for creating continuous smooth distributions of test norms. This paper describes two studies that explore the behavior of cubic splines in order to determine their appropriateness for use in test norming. The first study uses data from the Curriculum Referenced Tests of Mastery (CRTM) and employs two smoothing techniques in determining distributions of percentile ranks (PR). Method 1 calculated PR values from cumulative frequency counts of raw scores values and then smoothed them to create a continuous distribution. Method 2 smoothed the raw frequencies to create the continuous distribution. The second study was designed to generate a theoretical population, sample repeatedly from it, and smooth the results in a comparative study of the two procedures. Both studies showed that spline smoothing techniques are a convenient and economical procedure for analytically smoothing test score distributions and generally give good results. The second study found that, of the methods method 2 produced better results and was more robust with regard to the choice of smoothing parameter. Both methods were easy, efficient and economical to use. Statistical tables of standard deviations and means are included. (EGS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED247307

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

AN INVESTIGATION OF TWO PROCEDURES FOR SMOOTHING TEST NORMS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P. B. Jones

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Patricia B. Jones

Darrell L. Sabers

University of Arizona

Paper Presented to the 1984 Annual Meeting
American Educational Research Association
New Orleans, Louisiana

TM 840 190

AN INVESTIGATION OF TWO PROCEDURES FOR SMOOTHING TEST NORMS

Patricia B. Jones
Darrell L. Sabers
University of Arizona

Psychometricians have developed a variety of techniques for creating continuous, smooth distributions to norm and equate test data. Angoff(1971) and Kolen(1983) have reviewed some of these techniques. For example, hand smoothing permits direct control of the adherence of the function to the frequency distribution, but is not analytical in nature and is therefore ill-suited for computer applications. Traditional linear interpolation is useful for determining intermediate values of a function but is not helpful in eliminating irregularities caused by sampling error or test unreliability. Specific functions (e.g., logistic and quadratic) have been suggested by Marco(1977) and Lindsay and Prichard(1971) to fit sets of data. While these procedures do lend themselves to computer applications, one must assume that the functional form of the equation satisfactorily fits the data. Researchers familiar with item response theory are well aware that the exact form of the function used to fit a given set of data can engender substantial controversy.

If the functional form of the equation is not needed, a general analytical curve-fitting procedure permitting the creation of a continuous, smooth distribution is useful. Furthermore, it is desirable to be able to control the smoothness of that distribution by some easily specifiable parameter. Kolen(1983) has found cubic spline smoothing procedures to be effective in the smoothing of equipercentile equating functions. These procedures are very general, however, and can be used

to smooth a variety of functions regardless of the shape of the distribution. Spline techniques are used in fields as diverse as shipbuilding (Theilheimer and Starkweather, 1961) and astronomy (Jones, 1973, 1977) but have not been frequently applied in psychometrics. Two psychometric uses are found in their applications in multidimensional scaling (Winsberg and Ramsey, 1981) and test equating (Kolen, 1983). If one test is to be equated to another it may be necessary to create a continuous distribution in order to approximate ability levels that are not discrete. Furthermore, smoothing of score distributions may be desirable to approximate the distribution of the population from which scores are assumed to have been sampled. For these applications cubic spline curve fitting techniques appear to be well-suited.

Description of Cubic Smoothing Splines

For some time polynomial curves have been used to create continuous distributions for sets of discrete data. The primary drawback to such procedures is that a substantial number of terms must be incorporated to fit the obtained data points. As the number of these points increases, so in general does the order of the polynomial. Relatively small changes in ordinate values of data points can lead to substantial changes in polynomial coefficients. Moreover, the need to incorporate large-order polynomial terms leads to additional computational instability. The problems encountered in the use of polynomial approximations therefore tend to make them a less attractive solution to curve fitting than might be expected.

An alternative procedure outlined in Ahlberg, Nilson and Walsh (1967) is to connect piecewise polynomial segments such that many

short segments of some specified functional form are joined at locations determined by the analyst. This procedure gives a great deal more flexibility in curve fitting since one may connect points with line segments whose functional form is relatively simple while maintaining the ability to fit complex curves. The simplest example of this fitting technique is in the use of continuous piecewise linear line segments for the purpose of interpolation. However, piecewise linear functions are inaccurate for curvilinear situations, particularly in regions of substantial curvature such as might be expected near the extremes of cumulative frequency distributions.

Although these piecewise segments may take on any of a number of functional forms, in practice piecewise cubic polynomials are frequently used because they display certain optimal properties such as minimum curvature (Ahlberg et al., 1967) while retaining a relatively simple form. The minimum curvature property is predicted by a theorem of Holladay and is useful since it avoids the problems of oscillatory functions sometimes encountered in fitting a single polynomial to a large number of data points. The entire collection of piecewise functions is called a spline; the points at which the line segments are joined are called knots or ducks. In order to display the optimal properties described above, the piecewise segments are subjected to several restrictions, one of which is that the first and second derivatives must be continuous at the knots (i.e., the values of the first and second derivatives of the curve must be identical at the knot regardless of the spline segment on which they are evaluated).

DeBoor(1978) and Reinsch(1967) have discussed the use of smoothing splines. In these procedures a spline consisting of piecewise polynomials is fitted to the data as above but a certain amount of

latitude is permitted in the fit of the curve to the data. Whereas the procedures outlined by Ahlberg et al. restrict the ordinate value of the curve at the knots to the obtained data values, the fitted value is permitted to deviate from these values in smoothing splines. The amount of this departure is controlled by the specification of a smoothing parameter called E in this paper. A value of E=0 implies that the cubic spline is simply a spline of interpolation consisting of segments of cubic lines joined at the knots, the procedure outlined in Ahlberg et al. A larger value of E specifies the extent to which the function is allowed to deviate from the obtained data points. The limiting value of E is determined by the total deviation sum of squares of the data points from a straight line fitted to the data points and in its limiting case, the spline simply becomes a straight line.

The spline function S(x) is evaluated by determining the locations of the knots, x_j, representing score points in this paper. The value of the spline function for a given value x then represents the ordinate value of the distribution undergoing smoothing (i.e., percentile rank, observed frequency, etc.). Using the notation of Ahlberg et al. we have

$$S(x) = M_{j-1}\frac{(x_j - x)^3}{6\,h_j} + M_j\frac{(x - x_{j-1})^3}{6\,h_j} + \left(y_{j-1} - \frac{M_{j-1}\,h_j^2}{6}\right)\left(\frac{x_j - x}{h_j}\right) + \left(y_j - \frac{M_j\,h_j^2}{6}\right)\left(\frac{x - x_{j-1}}{h_j}\right) \quad (1)$$

where $h_j = x_j - x_{j-1}$, the length of the interval of interpolation; $x_{j-1} \leq x \leq x_j$; and the values of M_j are determined by solving the following equation:

$$\begin{bmatrix} 2 & \lambda_0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \mu_1 & 2 & \lambda_1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \mu_2 & 2 & \lambda_2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2 & \lambda_{N-2} & 0 \\ 0 & 0 & 0 & 0 & \cdots & \mu_{N-1} & 2 & \lambda_{N-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & \mu_N & 2 \end{bmatrix} \begin{bmatrix} M_0 \\ M_1 \\ M_2 \\ \vdots \\ M_{N-2} \\ M_{N-1} \\ M_N \end{bmatrix} = \begin{bmatrix} d_0 \\ d_1 \\ d_2 \\ \vdots \\ d_{N-2} \\ d_{N-1} \\ d_N \end{bmatrix} \qquad (2)$$

Here λ_j is defined as $\dfrac{h_{j+1}}{h_j + h_{j+1}} = \dfrac{x_{j+1} - x_j}{x_{j+1} - x_{j-1}}$,

μ_j is defined as $1 - \lambda_j = \dfrac{x_j - x_{j-1}}{x_{j+1} - x_{j-1}}$,

and d_j is defined as

$$6 \left\{ \frac{(y_{j+1} - y_j)/h_{j+1} - (y_j - y_{j-1})/h_j}{h_j + h_{j-1}} \right\}$$

The M_j are referred to as moments of the spline function and are second derivatives of the function evaluated at the knots. Since the x_j and y_j are known in this application, the entire system of equations may be solved through the use of well-developed, efficient recursion techniques.

While Holladay's theorem requires that $f(x_j) = y_j$ for all x_j in practice this restriction is not necessary. Instead one may impose some minimization requirement on the function; this can be done by requiring a least-squares solution for a spline with fewer knots than data points or by restricting $\sum_{j=1}^{n} [y_j - S(x_j)]^2 \leq E$. In this case the spline becomes the smoothing spline mentioned above.

The use of spline functions represents an analytical procedure which combines flexibility of form, replicability of curve-fitting techniques, and computational ease with a subjective element which permits the analyst to control the fit of the curve. Rather than prior

specification of E, several values in a range generally found to provide an good blend of curve fit and smoothness may be tried before choosing the value giving the best results. Comparing the smoothed curve and the original distribution on the same graph for several values of E may assist in the selection of E. Interpolated values of the spline function can be used for computer plotting within the same program which calculates the values of the smoothing function at the knots.

Determination of the optimal smoothing range is largely a matter of experience when the shape of the true distribution is unknown. Reinsch suggests restricting E to the interval $N - \sqrt{(2N)} \leq E \leq N + \sqrt{(2N)}$ where N represents the number of possible scores. For an 80 item test, this places E between 68 and 94. However, as Kolen (1983) points out, Reinsch assumed that the y_j were independent, an assumption not met when smoothing a curve of cumulative frequencies, for example. In practice, the optimal value of E depends on the shape of the curve being smoothed, the error within the sample (which is in part a function of the number of subjects sampled) and the size of the y_j. Optimal values of E will differ for cumulative frequencies and cumulative relative frequencies.

The evaluation of the spline function can be carried out through use of Fortran-callable subroutines available through IMSL(1982) or by use of deBoor's (1978) subroutines SMOOTH, SETUPQ, AND CHOL1D. The latter procedures were used in the analyses described in this paper. They are well-documented, appear to be quite robust, and easily modifiable by the user; furthermore they are quite economical and are easily transported from institution to institution.

In the present study smoothing splines were developed for

distributions of percentile ranks by two methods. In Method 1, the PR values were determined from cumulative frequency counts of raw score values and then smoothed, creating a continuous distribution of PR values. In Method 2, the raw frequencies were smoothed to create a continuous distribution. The area under the curve was then determined by integrating (1) to obtain the following:

$$\int_{x_{j-1}}^{x_0} S(x)\,dx = -\frac{M_{j-1}\,h_j^3}{24} - \frac{M_{j-1}(x_j - x_0)^4}{24\,h_j} + \frac{M_j\,(x_0 - x_{j-1})^4}{24\,h_j} + \frac{y_{j-1}\,h_j}{2}$$

$$- (y_{j-1} - \frac{M_{j-1}\,h_j^2}{6})\,(\frac{x_j - x_0}{2\,h_j})^2 + (y_j - \frac{M_j\,h_j^2}{6})(\frac{x_0 - x_{j-1}}{2\,h_j})^2 \qquad (3)$$

where $x_{j-1} \leq x_0 \leq x_j$ and M_j, M_{j-1}, h_j, x, and y are defined as before. When $x_0 = x_j$ this simply reduces to

$$\int_{x_{j-1}}^{x_j} S(x)\,dx = \frac{(y_j + y_{j-1})\,h_j}{2} - \frac{(M_j + M_{j-1})\,h_j^3}{24} \qquad (4)$$

This latter area is essentially the area of a trapezoid with a cubic correction for curvature. After integrating the entire curve, the area under the curve is normalized to 100. If 0 represents the lowest possible score, x_{max} the highest, and x any score on the interval $0 \leq x \leq x_{max}$, then

$$\frac{\int_0^x S(x)\,dx}{\int_0^{xmax} S(x)\,dx} \times 100$$

represents the PR for a subject with a score of x.

This paper represents an attempt to explore the behavior of cubic splines in order to determine their appropriateness for use in test norming. Two studies were carried out: the first explores the feasibility of the use of spline functions for data obtained in norming

the Curriculum Referenced Tests of Mastery (CRTM) (Sabers and Sabers, 1984). Because questions were raised concerning which of the two smoothing techniques more closely approximated the population from which the data were sampled, a second Monte Carlo Study was designed to generate a theoretical population, sample repeatedly from it, and smooth the results in a comparative study.

Study One

Methodology:

Initial explorations of the behavior of cubic spline functions were made using data from the Fall, 1982, administration of prototype mathematics and reading tests of the CRTM. The prototype tests were administered both on- and off-level to samples of 300 to 400 students per test. Grade levels of booklets and students ranged from first to eighth grade. Once the authors were familiar with the behavior of cubic splines, splines were used to smooth 40 distributions resulting from the Winter, 1983, administration of standard reading and mathematics tests on- and off-level to samples of 3000 to 4000 students.

The two smoothing procedures described previously were used on the prototype tests; appropriate values for the smoothing parameter E were determined for each distribution and procedure and the results compared for problems in curve fit. Because the criterion distribution is unknown, it is not possible to determine which method better approximates the true distribution of PR values.

Results and Discussion:

Results for Method 1 appeared quite good; small irregularities in PR distributions were effectively removed and with proper attention to the values of the smoothing parameter, smooth distributions were obtained. Values of E which appeared appropriate for the Fall prototype tests ranged between 10 and 50. Discernible differences in smoothed test score distributions were noted only for increments in E of five or more, so relatively few analyses were needed to obtain a satisfactory

distribution in most cases.

Little smoothing was required for the January norming of the st lard tests. With representative national samples of over 3000 students, PR distributions were quite smooth. For the sake of uniformity in reporting norms, all values of E were set to zero (i.e., no smoothing, interpolation only) and spline functions were used to determine intermediate values for subsequent test equating. Because CRTM standard test norms are often used as a basis for reporting norms on custom CRTM Tests, it is imperative that PRs be available for non-integer scores.

Two difficulties arose when using Method 1 for non-zero values of E. First, when several anomalously high or low frequencies appear close together, the spline function is ineffective in smoothing out the anomaly. Increasing values of E can cause the smoothing procedure to have minimal impact on regions needing smoothing while having greater impact on regions needing less smoothing. Second, end values of the spline function tended to depart from the asymptotic values of 0 and 100 with increasing values of E, a result of the tendency of the function to approach a least-squares line. This resulted in PR values greater than 100 or less than 0.

Method 2 was developed to solve the problems mentioned above. Since frequencies, unlike cumulative frequencies, represent independent observations, it was hoped that the impact of several anomalous successive values could be minimized. Although a slight improvement was noted in using Method 2 the problem was not totally solved. A potential solution to the problem lies in the use of variable length intervals between data points; i.e., binning data in such a way that the impact of too-large or too-small frequencies is minimized. Still another

potential solution lies in the use of least-squares splines rather than smoothing splines. Subsequent studies are planned to examine these procedures.

The second problem noted, the non-asymptotic ends of the distribution, was solved by the second procedure. Occasional negative smoothed frequencies were noted near the extremes of the distribution with the smaller samples in the Fall testing, but this is unlikely to occur when the number of subjects is large. It should also be noted that with small numbers of subjects and small values of E the spline function established in the either procedure is occasionally not monotonic.

Appropriate values of E for Method 2 were substantially larger than for the first procedure. Although appropriate values in Method 1 ranged between 10 and 50 for tests containing 50 to 80 items, the values for Method 2 ranged between 300 and 700. This is due in part to the substantial differences in the appearance of the curves being smoothed in each method and these differences are of no consequence in the evaluation of the two methods.

Both methods produced very good, usable results. Because the procedure is analytical, consistent results were obtained in smoothing data distributions while substantially reducing the time and expense of smoothing and interpolating distributions.

Study Two

Methodology:

While cubic splines were found to be an effective smoothing procedure in Study One and while Method 2 appeared to give slightly better results in smoothing test data, the goal in smoothing a distribution is to remove irregularities to approximate the test score distribution in the parent population. Since nothing is known about the parent population except that which can be generalized from a sample, it is not possible using student sample data to determine which of the two methods more closely approximates the actual distribution of test scores in the population. To answer this question, a Monte Carlo study was undertaken.

The following assumptions of true score theory were made:

1. $x = t + e$, the observed score is the sum of a true score and an error score.
2. $\sigma_x^2 = \sigma_t^2 + \sigma_e^2$, the observed score variance is the sum of true score variance and error score variance.
3. $\sigma_e^2 = (1 - \rho_{xx})\sigma_x^2$, error score variance is a function of observed score variance and test reliability ρ_{xx}.

Distributions of observed score values were simulated by sampling 400 pairs of random normal deviates. Setting $\mu_t = \mu_x = 36$ and $\sigma_t = 12$, and letting z_1 and z_2 represent the pair of random normal deviates, observed score values were computed according to the following formula:

$$x = (\sigma_t z_1 + \mu_t) + \frac{\sigma_t}{\sqrt{\rho_{xx}}} \sqrt{1 - \rho_{xx}}\ z_2$$

$$= (12 z_1 + 36) + \frac{12}{\sqrt{\rho_{xx}}} \sqrt{1 - \rho_{xx}}\ z_2$$

Item values were then grouped into discrete categories with a 1-point interval. The resulting distribution was then smoothed by the two methods described previously and smoothed PR (PR_s) values were compared with the PR values for a normal (PR_n) distribution with a mean of 36 and true-score standard deviation of 12 by calculating the value of T, the total fit error, as below:

$$T = \sum_{i=1}^{n} (PR_s - PR_n)^2, \qquad (5)$$

The smoothing procedure which yields the lowest value of T is presumed to give the closest fit for the set of data under consideration. While this is the procedure actually employed in this study, others could be used depending on the needs of the analyst.

The values of T are not directly comparable until optimal smoothing (minimizing the value of T) has been carried out for each smoothing procedure. To determine the value of E which minimizes T, IMSL subroutine ZXGSN was used. The method employs a golden section search procedure (IMSL, 1982) which assumes that a function has a unique minimum within the range of values under consideration. For smoothing procedures local minima rather than a unique minimum value may be encountered. The failure to find a unique minimum results from differing values of E causing a function to approximate more closely the criterion in one region while causing it to depart from the criterion in another. However, since T is compared for the two smoothing procedures, if the procedure which failed to converge produces a local minimum anywhere for which T is less than the procedure which converged, we may assume the procedure with the unknown absolute minimum yields better results for the given data set. In the present study,

Method 1 always converged to a minimum. Results for Method 2 were more variable. These results and their implications are discussed below.

Values of ρ_{xx} used in the study were 1.00 (no measurement error), 0.91 (reliability comparable with that of many well standardized achievement tests currently available), and 0.74 (permitting incorporation of substantial errors of measurement). For each value of ρ_{xx}, fifty sampling distributions were established. By specifying the same initial seed value, the distributions were restricted to the same sampled values across the three values of ρ_{xx}. Thus results for smoothing the first distribution at ρ_{xx} = 1.00 may be compared with smoothed results of the same distribution at ρ_{xx} = 0.91. Differing values of T in these distributions are due to the relative magnitude of measurement errors.

Results and Discussion:

Tables 1, 2, and 3 present the values of T and optimal smoothing values of E for the two procedures at ρ_{xx} = 1.00, 0.91 and 0.74 respectively. For comparison with T_1 and T_2, the summed squared deviations T_0 of the unsmoothed PR values from the PR values for a normal distribution are also presented in each table in column 6. These vary over the three distributions because of the relative contribution of measurement error to the procedure. Asterisks denote values of E which failed to converge to a minimum.

In every case with either Method 1 or Method 2, the smoothed result better approximates the given normal distribution than does the unsmoothed distribution indicating that the two analytical smoothing procedures are at least partially successful in recovering the distribution from which they were sampled. While in some cases the

improvement may not be substantial, in others overall deviations are markedly reduced. Figure 1 presents results of one such smoothing at ρ_{xx} = 1.00. Here deviations from the normal distribution for Method 1 are represented by a star; deviations for Method 2 by a triangle and deviations for the unsmoothed distribution by a solid line. The corresponding values for the sample (#41) for T_1, T_2, and T_0 representing the summed squared deviations are 20.89, 19.54 and 65.90 respectively. For this set of smoothed results, a marked improvement can be observed for either method and the effects of sampling error are clearly diminished.

For ρ_{xx} = 1.00 Method 2 yields better results than Method 1 in approximately half the cases although in general results from the two procedures are quite close(see Table 1). This indicates that the exact smoothing procedure chosen for a highly reliable test may be of less importance than the fact of smoothing itself. As the unreliability of the test increases, the superiority of Method 2 to Method 1 in terms of minimizing the error becomes more apparent. When ρ_{xx} = 0.91, Method 2 gives better results than Method 1 in 76% of the cases (see Table 2). Although results are generally fairly close (as in Sample #1 for which T_1 and T_2 equal 51.90 and 50.89 respectively), sometimes the discrepancy is greater (as in Sample #32 for which the corresponding values were 96.38 and 84.40). When ρ_{xx} = 0.74 the discrepancy between the two procedures becomes more substantial (see Table 3). Method 2 is clearly superior to Method 1 in all but four of the cases. In three of these four, convergence to an optimal smoothing parameter did not occur (in fact, it did not occur in most cases of Method 2 at this reliability). With choice of another minimization procedure it is

possible that lower values of $T_2 < T_1$ would be found for at least some of these values.

One effect which stands out quite sharply is that as test reliability decreases, the smoothing procedure loses its effectiveness in recovering the original normal distribution. Comparison of distributions for a given run over the three reliabilities shows that, as expected, the net effect of measurement error is a flattening of the entire distribution of test scores yielding a more platykurtic distribution. At the same time, some of the larger sampling anomalies are smoothed out. Therefore, measurement error in itself has a smoothing effect; furthermore, once the distribution systematically departs from the original, as happens when the distribution is flattened, both smoothing procedures lose their efficacy in recovering the original distribution. For this reason, spline smoothing procedures cannot be expected to compensate for poor test reliability.

Means and standard deviations of T_1, T_2, T_0, E_1, and E_2 are displayed in Table 4. Several conclusions are readily apparent in these tables.

1. While average sizes of T_1 and T_2 are close for $\rho_{xx} = 1.00$, values diverge as ρ_{xx} decreases with $\overline{T}_2 < \overline{T}_1$.
2. The average size of the smoothing parameter tends to decrease as reliability decreases regardless of the smoothing procedure chosen.
3. As reliability decreases, the use of cubic spline smoothing procedures loses its effectiveness in restoring the shape of the original distribution.
4. The variances of T_1, T_2, and T_0 increase as reliability decreases.

5. The variances of E_1 and E_2 decrease as reliability decreases.

These results, therefore, confirm the conclusions based on examination of results of individual runs presented in Tables 1, 2, and 3.

The failure of the convergence procedure for Method 2 in most cases when $\rho_{xx} = 0.74$ should not be considered a weakness, but rather a strength. It can be noted that despite the failure of the minimization procedure to find an optimal smoothing parameter, Method 2 generally gave superior results for the value of E selected. In fact, there exists a range of smoothing values which can be expected to yield good results. It should also be noted that failure to converge is a weakness of the golden section search method, not the smoothing procedure. With test data, the optimization procedure used in this section could not and would not be used.

Conclusions

Both studies showed that spline smoothing techniques are a convenient and economical procedure for analytically smoothing test score distributions and generally give good results. Method 2, the method of integration, was developed in response to two problems encountered in Method 1-- the occurrence of several successively low or high values on the curve which proved resistant to smoothing, and the tendency of the ends of the distribution to depart from the horizontal. Method 2 improved the first problem and essentially eliminated the second.

In the Monte Carlo study it was found that both smoothing procedures give similar results for highly reliable tests ($\rho_{xx} = 1.00$) but as measurement error increases, Method 2 is found to give better results. Neither procedure was completely effective, however, in restoring the shape of the original distribution once measurement error had systematically changed the curve.

Optimal smoothing parameters tended to vary substantially from sample to sample. However, when a test is highly reliable, a larger smoothing parameter may well be necessary because the measurement error is not present to smooth out the effects of sampling error. Generally, Method 2 is quite robust with respect to the choice of smoothing parameter. For Method 1 the choice may be more critical and confined to a rather narrow range of values.

Because Method 2 generally gave results as good as or superior to Method 1 and because it was more robust with regard to the choice of smoothing parameter, Method 2 should be used whenever appropriate.

However, either method may be expected to give good results, particularly when tests are reliable, and both are easy, efficient, and economical to use.

Bibliography

Ahlberg, J.H., Nilson, E.N. and Walsh, J.L. The Theory of Splines and their Applications. New York, Academic Press, 1967.

Angoff, W.H. Scales, Norms, and Equivalent Scores. R.L Thorndike (ed.), Educational Measurement (2nd Ed.). Washington, D.C., American Council on Education, 1971.

de Boor, C. A Practical Guide to Splines. New York, Springer-Verlag, 1978.

International Mathematical and Statistical Libraries (IMSL). Reference Manual (9th Ed.), 1982.

Jones, H.P., 1973. The Formation of Resonance Lines in Multidimensional Media. III. Interpolation Functions, Accuracy and Stability. Astrophysical Journal, 185, 183-196.

Jones, H.P., 1977. Splines under Tension in Integral Transfer Problems. J. Quant. Spectrosc. Radiat. Transfer, 17, 765-776.

Kolen, M.J., 1983. Effectiveness of Analytic Smoothing in Equipercentile Equating. ACT Technical Bulletin #41. American College Testing Program. Iowa City, Iowa.

Lindsay, C.A. and Prichard, M.A., 1971. An Analytic Procedure for the Equipercentile Method of Equating Tests. J. Ed. Meas., 8, 203-207.

Marco, G.L., 1977. Use of the Logistic Model as an Alternative to Linear Interpolation for Computing Percentile Ranks. J. Ed. Meas. 14, 271-275.

Reinsch, C.H., 1967. Smoothing by Spline Functions. Numerische Mathematik, 10, 177-183.

Sabers, D.L. and Sabers, D. Curriculum Referenced Tests of Mastery. Charles E. Merrill Publishing Company, Columbus, Ohio, 1984.

Theilheimer, F., and Starkweather, W., 1961. The Fairing of Shiplines on a High-Speed Computer. Numerical Tables Aids Computation, 15, 338-355.

Winsberg, S. and Ramsey, J.O., 1981. Analysis of Pairwise Preference Data using Integrated B-Splines. Psychometrika, 46, 171-186.

Table 1. Deviations from Normal Distribution for Each Smoothing Method and Comparison Deviations for No Smoothing (T_1, T_2, and T_0); Optimal Smoothing Parameter Values (E_1 and E_2). ρ_{xx} = 1.00.

	T_1	E_1	T_2	E_2	T_0
1	83.75	18.56	84.54	461.85	105.53
2	85.98	6.31	88.12	206.77	94.73
3	112.66	29.77	117.61	329.14	146.42
4	26.56	6.54	22.27	265.52	32.39
5	305.63	16.25	306.46	330.18	328.16
6	282.62	32.78	269.72	369.00	319.79
7	50.16	6.23	51.02	341.52	59.48
8	167.96	21.48	173.01	259.08	197.45
9	538.01	87.78	625.96	260.28*	647.39
10	720.60	71.41	715.21	344.73	798.64
11	203.79	18.76	224.61	173.89*	225.45
12	383.01	64.68	373.85	314.33	460.87
13	581.71	117.84	563.12	382.20	704.10
14	12.57	11.65	12.86	366.16	25.76
15	479.56	62.66	471.28	334.32	553.58
16	196.77	9.66	195.11	324.79	209.80
17	149.83	3.54	145.86	97.09*	160.52
18	189.95	22.54	192.67	303.80	217.67
19	230.57	35.65	234.06	460.04	267.02
20	264.44	13.68	258.92	85.25*	282.96
21	4.66	31.32	3.12	493.78	40.02
22	196.77	38.00	191.07	454.74	241.31
23	40.49	45.94	35.73	253.52	93.04
24	88.50	8.75	86.90	130.42*	100.19
25	183.80	5.83	183.89	117.07*	191.39
26	100.47	10.24	94.18	333.62	113.67
27	151.44	33.01	153.97	517.84	184.57
28	150.90	11.32	145.74	306.56	164.09
29	411.65	115.18	400.84	285.33	530.89
30	87.04	44.80	84.93	294.82	131.99
31	68.54	13.44	63.00	366.78	84.65
32	34.90	20.39	36.75	445.52	58.17
33	104.60	10.15	101.54	310.07	118.87
34	29.50	18.67	30.42	224.30	54.57
35	340.82	21.39	349.51	21.22*	364.06
36	334.40	37.33	341.98	372.78	373.82
37	259.01	8.52	360.12	131.10	270.23
38	74.07	20.81	74.49	357.29	100.32
39	248.43	33.92	249.78	304.07	289.81
40	207.40	48.75	210.13	401.55	259.59
41	20.89	42.07	19.54	495.03	65.90
42	245.84	8.66	219.08	29.44	257.57
43	352.26	29.40	363.99	137.30	392.97
44	50.28	39.76	51.58	486.80	92.84
45	46.15	21.96	43.35	332.14	74.24
46	110.31	29.70	108.68	375.09	140.84
47	80.92	42.60	78.54	295.38	131.92
48	30.47	13.96	30.64	431.69	47.02
49	80.50	7.88	73.32	261.90	90.37
50	102.99	13.33	97.06	275.02	122.49

* indicates value of E_2 which failed to converge to a minimum.

Table 2. Deviations from Normal Distribution for Each Smoothing Method and Comparison Deviations for No Smoothing (T_1, T_2, and T_0); Optimal Smoothing Parameter Values (E_1 and E_2). $\rho_{xx} = 0.91$.

	T_1	E_1	T_2	E_2	T_0
1	51.90	33.83	50.89	506.88	93.59
2	176.72	6.99	174.74	278.85	185.87
3	94.25	10.11	82.16	251.77	106.47
4	254.79	7.42	255.11	361.80	265.12
5	279.74	18.06	276.85	434.73	305.24
6	297.61	36.41	283.07	497.51	342.73
7	35.68	23.22	33.49	331.14	69.14
8	255.40	13.25	251.22	296.71	272.55
9	437.60	13.01	403.17	46.22*	455.01
10	374.19	20.31	386.97	231.98*	398.48
11	118.24	12.18	116.55	557.14	131.78
12	110.48	24.93	108.36	322.78	145.04
13	239.61	35.57	239.27	332.31	282.92
14	94.15	7.55	93.03	324.60	104.24
15	181.83	10.68	175.97	257.54	195.79
16	271.59	14.67	272.87	358.98	291.38
17	83.49	17.52	78.28	314.29	107.49
18	142.40	14.24	138.64	332.05	159.53
19	638.41	18.28	648.31	335.36	660.03
20	204.67	25.28	194.57	490.34	239.48
21	99.23	13.42	105.03	83.70*	115.28
22	129.05	12.66	123.45	330.32	145.77
23	134.43	43.57	134.56	457.45	191.26
24	179.98	7.53	174.31	331.49*	189.87
25	492.85	6.99	471.34	190.41*	502.20
26	124.28	9.21	121.27	225.11	136.47
27	46.69	13.67	45.30	267.55	61.92
28	348.73	23.91	337.90	382.28	376.55
29	241.15	16.20	237.87	342.18	258.99
30	6.26	20.75	4.63	256.42	28.22
31	176.73	3.39	163.12	25.87	180.70
32	96.38	13.25	84.40	332.00	114.54
33	473.62	6.97	468.61	198.33*	482.66
34	80.91	9.87	75.40	193.99	93.63
35	263.61	7.60	255.47	95.99	273.83
36	1089.95	42.32	1092.79	475.63	1141.47
37	158.10	3.99	157.23	79.36*	162.35
38	53.68	7.97	50.68	314.19	63.15
39	441.10	11.96	442.34	409.03	455.24
40	98.56	25.09	100.00	390.84	130.70
41	147.61	25.70	144.00	322.74	182.59
42	482.16	7.80	479.04	206.96*	491.73
43	222.93	24.20	214.52	398.75	256.55
44	126.83	24.97	125.35	343.91	160.80
45	260.78	9.26	242.54	51.26	273.13
46	12.27	8.61	12.88	256.12	22.17
47	263.33	34.78	265.58	358.89	308.88
48	110.57	9.74	103.24	403.06	125.26
49	163.16	7.65	148.59	310.80	173.00
50	485.48	9.94	488.85	383.97	497.77

* indicates value of E_2 which failed to converge to a minimum.

Table 3. Deviations from Normal Distribution for Each Smoothing Method and Comparison Deviations for No Smoothing (T_1, T_2, and T_0); Optimal Smoothing Parameter Values (E_1 and E_2). $\rho_{xx} = 0.74$.

	T_1	E_1	T_2	E_2	T_0
1	241.66	10.94	220.86	212.36	254.65
2	548.42	8.80	498.20	138.34*	560.53
3	635.44	7.21	509.24	34.61*	645.81
4	993.73	9.54	943.37	58.29*	1005.17
5	913.36	7.73	872.68	183.86*	925.36
6	791.53	8.70	702.23	115.47*	802.89
7	242.01	11.16	200.21	71.42*	255.47
8	877.74	25.00	834.73	71.91*	911.00
9	1170.78	16.49	1067.33	58.04*	1193.64
10	610.88	5.05	577.59	39.79*	617.05
11	504.86	4.42	479.76	15.36	509.55
12	305.66	8.47	289.06	111.19*	318.00
13	391.18	8.77	377.93	346.33	404.82
14	484.15	2.67	487.68	36.60*	487.91
15	467.11	8.09	384.54	9.53*	478.60
16	849.73	6.00	852.49	143.63*	857.29
17	365.89	23.59	348.40	476.38	397.76
18	460.06	13.03	413.49	25.08*	479.10
19	960.47	13.53	960.93	404.00	977.07
20	568.97	15.53	484.15	43.95*	590.65
21	348.88	5.15	303.43	74.93*	354.70
22	567.28	7.20	532.89	37.00*	576.01
23	447.15	27.70	429.68	190.73*	485.80
24	713.23	6.61	654.23	37.68*	718.67
25	970.64	4.73	823.31	63.94*	976.56
26	499.23	14.00	494.92	174.17*	518.57
27	338.95	8.35	311.61	43.29*	350.99
28	535.10	11.33	481.71	227.23*	550.85
29	597.27	13.55	531.90	52.52*	615.38
30	164.18	8.87	158.55	272.45	175.91
31	814.95	3.80	725.76	129.74	819.50
32	563.62	6.82	487.20	102.98*	570.83
33	1131.19	8.38	1012.76	59.87*	1141.39
34	465.82	10.24	440.02	171.39*	480.60
35	818.65	4.39	762.21	98.17*	823.86
36	1904.61	21.94	1878.25	223.99*	1938.69
37	467.34	11.00	479.25	188.76*	484.26
38	308.66	7.12	290.95	292.54*	318.08
39	797.37	9.70	763.84	39.78*	811.00
40	442.70	5.10	399.21	64.82*	448.82
41	624.32	10.96	553.88	67.40*	639.14
42	972.38	3.87	889.62	12.41*	976.71
43	562.88	4.66	497.60	105.67*	567.99
44	512.46	4.28	469.98	97.73*	517.56
45	828.46	3.98	744.60	97.54*	833.17
46	340.77	8.17	296.13	67.00*	351.24
47	442.19	14.91	411.16	45.92*	462.91
48	574.83	4.49	497.68	117.21*	579.76
49	641.65	2.67	509.34	53.12*	642.94
50	1304.74	8.02	1233.36	12.44*	1316.71

* indicates value of E_2 which failed to converge to a minimum.

Table 4. Means and Standard Deviations of T_1, T_2, and T_0 (Summed Squared Errors) and E_1 and E_2 (Smoothing Parameter Values) for Each Reliability.

		T_1	T_2	T_0	E_1	E_2
ρ_{xx} = 1.00						
	Mean	186.1	186.2	220.4	24.8	307.0
	S D	159.8	163.0	179.7	25.7	122.5
ρ_{xx} = 0.91						
	Mean	227.1	222.7	248.2	16.5	305.6
	S D	190.7	191.0	193.0	10.1	122.5
ρ_{xx} = 0.74						
	Mean	641.7	591.4	654.4	9.5	116.4
	S D	315.3	304.6	316.7	5.7	102.7

n = 50

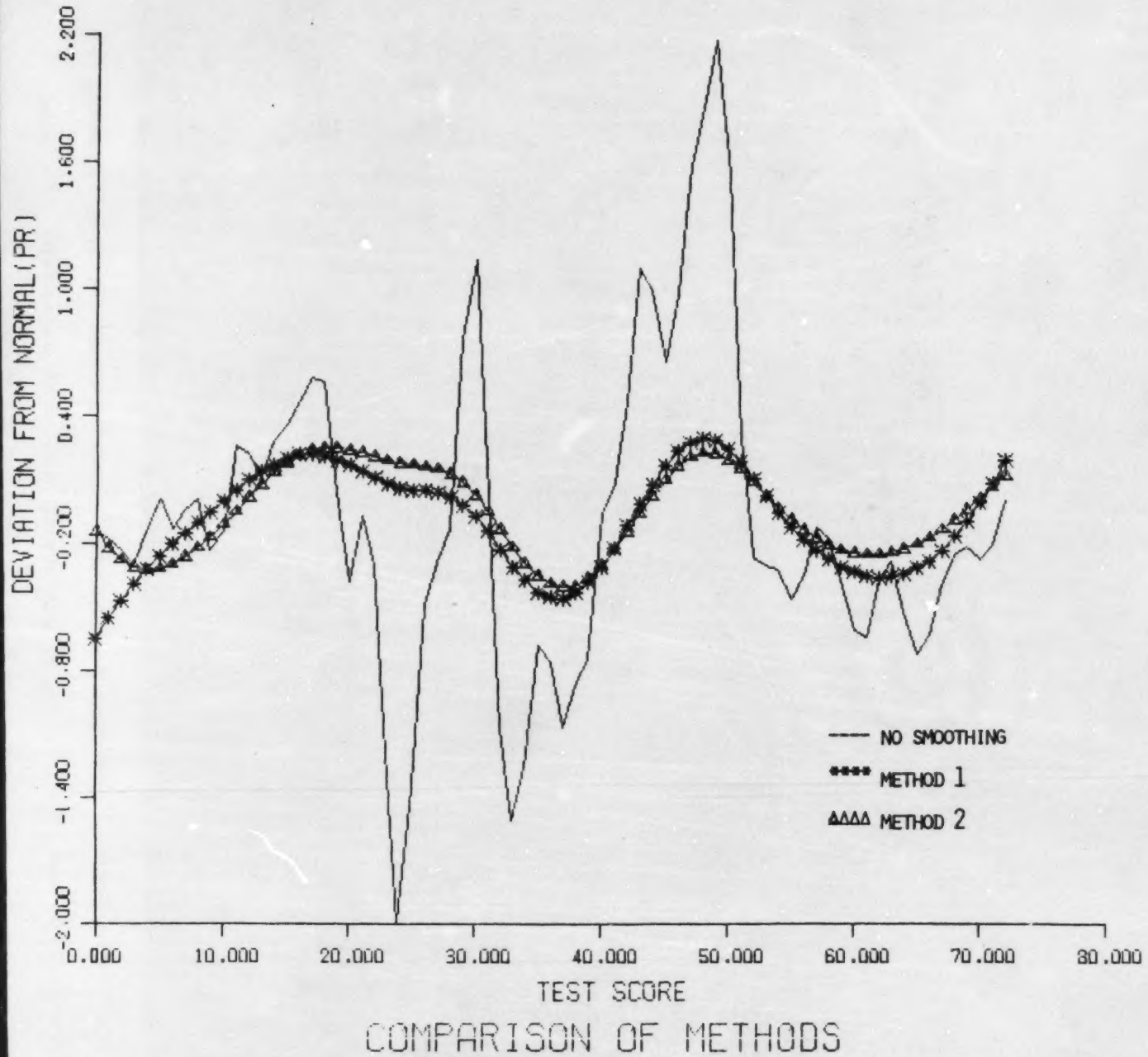

FIGURE 1. COMPUTER PLOT OF DEVIATIONS FROM NORMAL DISTRIBUTION OF PR VALUES. ρ_{xx} = 1.00.